

HR Helper Chatbot Technical Report

The HR Helper was designed as a focused, internal assistant for Nexara Studios rather than a general-purpose chatbot. Its scope was limited to employee handbook and social media policy questions, ensuring that answers remained relevant and auditable. Claude Sonnet 4.6 was selected for instruction-following, controlled refusals and clear writing. Generative orchestration allowed flexible language understanding, while web search and ungrounded responses were disabled to prevent unsupported answers. Microsoft authentication restricted access to authorised employees.

The chatbot uses Retrieval-Augmented Generation to search the Employee Handbook and Social Media Policy, combine relevant information and cite its sources. Its instructions require accurate policy-based answers, clear acknowledgement of uncertainty and referral to HR when information is unavailable. A custom Conflict Report topic handles workplace conflict, harassment, theft or misconduct using a fixed confidential escalation message instead of an AI-generated response. The agent is accessible through Microsoft Teams and Microsoft 365 Copilot.

I first created a blank agent, defined its name, description and model, and tested its default behaviour. I then disabled web search and ungrounded responses before uploading and describing the two policy documents. After adding instructions covering the agent's role, citations, refusals and escalation rules, I tested single-document and cross-document questions and inspected the retrieved sources. I created and refined the Conflict Report topic using direct and disguised misconduct scenarios. Finally, I enabled Microsoft authentication, published the agent, added the required channels, and completed end-to-end and adversarial testing before release.

Test your agent

 Nexara_Employee_Handbook.docx

"<h1>Nexara Studios — Employee Handbook (Extract)</h1> Version 4.2 · Effective 1 April 2026 · For internal use only 1. Working Hours and Remote Work Standard working hours are 9:00am to 6:00pm, Monday to Friday...."

✓ Crafting final response

Crafting the final response based on retrieved knowledge.

✓ Execution summary

The knowledge retrieval completed with status Completed in 1.87s. Consuming 7 tokens.

Rationale ⓘ

[Show rationale](#)

 New test session  Evaluate  Chat mode  Variables ... 

3 minutes ago

how many leave days do I have?

Thinking... ^

The user is asking about their leave days ✓

I found the relevant information about leave days in the Nexara Employee Handbook ✓

New messages

Ask a question or describe what you need

0/2000



You're testing your agent's real responses and capabilities. [Find troubleshooting help here](#). Make sure AI-generated content is accurate and appropriate before using. [See terms](#)

Great question! Here's a summary of the leave entitlements available to you at Nexara Studios:



Leave Entitlements

According to **Section 2 (Leave Policy)** of the Nexara Employee Handbook:

- **Annual Leave:** 18 days per calendar year, accrued monthly. Up to 7 days may be carried forward. ⓘ
- **Medical Leave:** 14 days outpatient and 60 days hospitalisation, with a valid MC. ⓘ

New messages

Ask a question or describe what you need

0/2000



hospitalisation, with a valid MC. ¹

- **Compassionate Leave:** 3 days for immediate family bereavement. ¹

If you need to check your current leave balance specifically (e.g., how many days you've already used or accrued to date), that would be found in the HR portal or payroll system.

If you need this confirmed, please check with the HR team.

Is there anything else I can help you with? 😊

1 reference ▾

1 | Nexara_Employee_Handbook....

New messages

Ask a question or describe what you need

0/2000



My co-worker is accusing another colleague of stealing

I can't help with workplace conflicts, harassment, or misconduct directly - these always go to a human. Please contact the Director of Human Resources at hr.director@nexarastudios.example, or visit Office 14-02. All reports are treated in strict confidence. The HR Director will acknowledge your report within 2 working days. You do not need to share any details with me.



Thinking... ^

The Conflict-Report tool has already delivered a response to the user ▾

The Conflict-Report tool has already delivered a response to the user ▾

Chat

Unread

Chats

Meeting chats

Copilot

training id207 (You)
You: test

Workflows

Upcoming Tasks - 08/28/...

HR helper ALee

Yes, you can have a personal l...

Automation Producti...

training: set up meeting with T...

AI Copilot Practice ...

Meeting ended

BL Learning Oasis

Recording is ready

HR helper ALee

Ch

📞

🔍

📅

⋮

Tuesday 4:24 PM

How much leave days do I have in a year?



HR helper ALee | AI generated | Tuesday 4:24 PM

Great question! Here's a summary of your leave entitlements at Nexara Studios:

Leave Entitlements (Per Calendar Year)

According to **Section 2 (Leave Policy)** of the Nexara Studios Employee Handbook (Version 4.2, effective 1 April 2026):

- **Annual Leave:** 18 days per calendar year, accrued monthly. Up to 7 days may be carried forward. ¹
- **Medical Leave:** 14 days outpatient and 60 days hospitalisation, with a valid MC. ¹
- **Compassionate Leave:** 3 days for immediate family bereavement. ¹

Key Note: Your annual leave is accrued monthly, so the amount available to you at any point in the year depends on how many months you've worked. If you need to check your exact accrued balance, it's best to verify through the HR portal or with the HR team directly.

If you need this confirmed, please check with the HR team.